

A)

	Motif 1	Motif 2	Motif 3	Motif 4
RV2837c_Mt	VGWCHVHPADTIGAGLALALVLDGCG [35]	VDLVVTWIPSVDRLGALG [6]	RELVVDHFEASND [10]	SADSTTTVAEILDANGKPIDPRVAHCIYAGLATDTGSRWASV
sll1253_Sep	DLILCHOTAFQVGLGARAVGLAKLHSGSR [34]	IRSLYIVNQGGDRLGKAA [8]	ROVAIYDHELNSP [11]	AVGASTTLIVEKQORADISLUMVEASVMALGIHVDTGSLTFTYT
MGA_Mg	IVIFHVRFPDGLGAGQGLFHLIKANF [32]	EALAIWPNYKNRIELRE [7]	KAVLRDHPHNEB [11]	SYVACCEQIVEMATVAKWTIPPVATLLYIGIYDTSNRFLYSNT
YTQI_Bs	IILHRHVRFPDAGSQCGLTLEILRETY [30]	GALVIVCPITANQERIDDDQR [4]	AKLMKDHPHNEB [13]	SVSEMIYELYLEGKEHGWKLNKTAELIYAGIVGDTGRFLFPNT
AF2029_Af	LGIFTHDNPPDMSMSAYALREIAKQFD [37]	YDAFAIVSSGPGVNNISIP [3]	DISIVDHPHNEB [10]	DVGATATILTEYIKELKITPSKILATLALFFGIKSEIDEFKRNR
MJ0977_Mj	NKILIVTHIDPGLTSRAILQKLABRLN [23]	YDLIHFAPLGGGQKLMKE [11]	DKIITDHPHNEB [18]	GAEICGAGVSYPAKAINNDWIDLAKYAVLGAAGDIQNIIEGKLI
MJ1198_Mj	RPIIRHHAQDGYCGGIALEKAILPII [46]	LPLIVDNGSTDEDIPAI [7]	IEVIVDHPHNEB [85]	KGRTYDREYLEKIALCWDFAFYLRFRMDGKGVDPILATNIKEP
HP1042_Hp	MOVYHLSHIDGYACQLVSKQFFKNIQ [27]	EFLIIVSPLNENINEAEYL [13]	IQIQLDHPHNEB [19]	IVYEFLKHYAILEPKNTTWLEPLVEMVNSVDIADTQGYGFELG
RecJ_Hi	QKIVIVGDFDAGATSTALSVLALRQLG [31]	VQLIMVNGVSSFDGVAE [5]	IRVLVDHPHNEB [33]	LAVRAKPRELGIPTAETQPNFTDLDLVALGTIADVVPDQNNR
RecJ_Hp	TEILVVGDFDAGGVISSAIMAKFFESLN [27]	APLIITVAGINAFEAARF [5]	YTLIITDHECLHH [27]	LVAFYLCYGIHQLLGKEKSHSSELLCLAGVATIDMPLTFFNR
RecJ_Sep	EKVITINGDFDAGITSTAVLWEGIGQFF [32]	TKLIIVTPTGSTNLDEIVY [5]	MDVIVDHPHNEB [27]	VAFKLVEALYNQYPTVPQQPLEDLDLVAIGLIADLVTLQGDOR
YYBQ_Bs	ILIEGHQNFETITICSAIAYADLKNKLG [36]	VNGVILVCHNERQQSIKDI [3]	QVLEVDHHERIAN [12]	PVGCTATILNKMYKNNVKIEKEIAGLMLSAIISPSLLPKSPTC
ICRA_Sg	ILVEGHQNFESPAIGSSYAFAYLAREAY [38]	AEQVILVCHNEFOQSADI [3]	EVYGVNDHHRVAN [12]	PVGSASSIVYRMFKHSVAVSKEIAGLMLSGLSITLLLSKSPPT
MJ0608_Mj	RYVGHKNFETPSIASAIVLAYFLDCYP [31]	GKEITLVCHSEKSSQSFDDL [3]	KLIATDHPHNEB [19]	IAELYFKDAIDLIGGKKKELKPDLAGLLLSAITSPTVLPKSPPT
AF0756_Af	VTVGHKNFETPSVCSAIAFAYLWKNWK [46]	GKKVAVLCHSEKAQTVDGI [3]	EWVAIVDHPHNEB [12]	PVGCTATVIKLLFDKTVGEIPKDIAGILLSSLSITVLPKSAAT
U60409_Lm	TVVQNGEGDPSIVGCIYLAFLPDQKP [44]	QIAHNDVLAALNASVVLY [14]	RVVGVDHPHNEB [11]	LRTVGSACTLVTELYRECGEDVVCPTLLTAPIVLDVWPEPAQK
PPX1_Sc	TICVGNESADPSIASAITYSYCOYIYN [62]	ELNSYLVNNDTPKLNKNY [2]	NVVGVDHPHNEB [14]	SCSSLVFNVYVEKLGQDREVVMNIAPLIMGAILLIDTSNMRKVB
PRUNE_Dm	HLVGMNESOGLSAVSATVLAFLVYQQRH [48]	DVNVLVNDHHSPLAPNVT [1]	NVTEFIDHPHNEB [18]	SVGSCATLVAQRYLAEDQPRSTSVQAQLLHATIVLDITINPAPAAK
h-prune.1_Hs	HVLGNEACGLSTVSALALAFYLAKTT [52]	QUTLILVCHHILSKSDTAL [2]	AAVEVIDHPHNEB [12]	LVGSCATLVTERILOGAPEILDRTAALLHGTILDCVNMDLKIG

	Specific motifs	Gis	Family 1	Family 2
RV2837c_Mt	[107] TVNLAASGFGCGGHRLAAGYTTGS	1648883		
sll1253_Sep	[123] DTDLTQLLEPYGGGHAQAANLNDV	1653244		
MGA_Mg	[102] GINVRDIAIKYGGGHNASGAIITNK	1045875		
YTQI_Bs	[103] GPVINGLARKYNGGHPPLASGASTYSW	2293259		
AF2029_Af	[102] EVLRRAFDVGSGAGGHAHAAGAIPLG	2648507		
MJ0977_Mj	[254] AIKYASEKVNCGGKHKFAAGAYIPDN	2128614		
MJ1198_Mj	[109] QLMEEIPEASLDGGGHECAGSLKPFVEG	2128720		
HP1042_Hp	[137] CDVCELSQMCFCNGGHRNASGGKIDGP	2314198		
RecJ_Hi	[168] RIHSQHPNMILKFGGHAMAGLSIRBE	1172895		
RecJ_Hp	[155] DALNGVSSLLLGYGHRQACGLSVEKN	2313437		
RecJ_Sep	[165] ALLHSQRHMLGFGGHPFAAGLSLPLD	1652638		
YYBQ_Bs	[31] DLSRKTVBELISLDAKEFTIG [75]	TALLKGVVSRKKQVVPVLT	586817	
ICRA_Sg	[31] NLASKSAEELIDIDAKTFELN [75]	HAPLAGAVSRKKQVVPQLTE	1743856	
MJ0608_Mj	[32] VVGKLKPEEITINMDFKNFDFN [74]	SVFEGVMSRKKQVVPPLER	1591318	
AF0756_Af	[32] AVDDLTAAMDIAKDYKDFDMS [75]	SVNLGVMRKKQVVPPLEK	11498362	
U60409_Lm	[36] DVLALSVPQILRRDYKQFSFK [105]	YSLSDPSISRKK-LVPALSE	1407725	
PPX1_Sc	[41] DIKGFVSVDILKDYKQFNQ [97]	MFKQLNVEATRKQVVPYLEE	730369	
PRUNE_Dm	[37] DISKLTLEVLKDKMKVLQTD [97]	LRQHNVAQTRKH-ILPIVKR	1079081	
h-prune.1_Hs	[34] DVSGLTTEQMLRKDKQKTYRQ [94]	YLOGNTQVSRKK-LPLLLQE	11245938	

Fig. 1

24308203